

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Antelman, Douglas  
Gregory, Richard J.  
Wils, Kenneth N.
- (ii) TITLE OF INVENTION: Tissue Specific Expression of  
Retinoblastoma Protein
- (iii) NUMBER OF SEQUENCES: 46
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
  - (B) STREET: Two Embarcadero Center, 8th Floor
  - (C) CITY: San Francisco
  - (D) STATE: CA
  - (E) COUNTRY: USA
  - (F) ZIP: 94111
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/801,092
  - (B) FILING DATE: 14-FEB-1997
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/751,517
  - (B) FILING DATE: 15-NOV-1996
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Fitts, Renee A.
  - (B) REGISTRATION NUMBER: 35,136
  - (C) REFERENCE/DOCKET NUMBER: 016930-001020
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 415-576-0200
  - (B) TELEFAX: 703-576-0300

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 437 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met 1	Ala	Leu	Ala	Gly 5	Ala	Pro	Ala	Gly	Gly 10	Pro	Cys	Ala	Pro	Ala	Leu 15
Glu	Ala	Leu	Leu 20	Gly	Ala	Gly	Ala	Leu 25	Arg	Leu	Leu	Asp 30	Ser	Ser	Gln
Ile	Val	Ile 35	Ile	Ser	Ala	Ala	Gln 40	Asp	Ala	Ser	Ala	Pro 45	Pro	Ala	Pro
Thr	Gly 50	Pro	Ala	Ala	Pro	Ala 55	Ala	Gly	Pro	Cys	Asp 60	Pro	Asp	Leu	Leu
Leu 65	Phe	Ala	Thr	Pro	Gln 70	Ala	Pro	Arg	Pro	Thr 75	Pro	Ser	Ala	Pro	Arg 80
Pro	Ala	Leu	Gly	Arg 85	Pro	Pro	Val	Lys	Arg 90	Arg	Leu	Asp	Leu	Glu 95	Thr
Asp	His	Gln	Tyr 100	Leu	Ala	Glu	Ser	Ser 105	Gly	Pro	Ala	Arg	Gly 110	Arg	Gly
Arg	His	Pro 115	Gly	Lys	Gly	Val	Lys 120	Ser	Pro	Gly	Glu	Lys 125	Ser	Arg	Tyr
Glu	Thr 130	Ser	Leu	Asn	Leu	Thr 135	Thr	Lys	Arg	Phe	Leu 140	Glu	Leu	Leu	Ser
His 145	Ser	Ala	Asp	Gly	Val 150	Val	Asp	Leu	Asn	Trp 155	Ala	Ala	Glu	Val	Leu 160
Lys	Val	Gln	Lys 165	Arg	Arg	Ile	Tyr	Asp	Ile 170	Thr	Asn	Val	Leu	Glu 175	Gly
Ile	Gln	Leu	Ile 180	Ala	Lys	Lys	Ser	Lys 185	Asn	His	Ile	Gln	Trp 190	Leu	Gly
Ser	His	Thr 195	Thr	Val	Gly	Val	Gly 200	Gly	Arg	Leu	Glu	Gly 205	Leu	Thr	Gln
Asp	Leu 210	Arg	Gln	Leu	Gln	Glu 215	Ser	Glu	Gln	Gln	Leu 220	Asp	His	Leu	Met
Asn 225	Ile	Cys	Thr	Thr	Gln 230	Leu	Arg	Leu	Leu	Ser 235	Glu	Asp	Thr	Asp	Ser 240
Gln	Arg	Leu	Ala	Tyr 245	Val	Thr	Cys	Gln	Asp 250	Leu	Arg	Ser	Ile	Ala 255	Asp
Pro	Ala	Glu	Gln 260	Met	Val	Met	Val	Ile 265	Lys	Ala	Pro	Pro	Glu 270	Thr	Gln
Leu	Gln	Ala 275	Val	Asp	Ser	Ser	Glu 280	Asn	Phe	Gln	Ile	Ser 285	Leu	Lys	Ser
Lys	Gln 290	Gly	Pro	Ile	Asp	Val 295	Phe	Leu	Cys	Pro	Glu 300	Glu	Thr	Val	Gly
Gly 305	Ile	Ser	Pro	Gly	Lys 310	Thr	Pro	Ser	Gln	Glu 315	Val	Thr	Ser	Glu	Glu 320
Glu	Asn	Arg	Ala	Thr 325	Asp	Ser	Ala	Thr	Ile 330	Val	Ser	Pro	Pro	Pro 335	Ser

Ser Pro Pro Ser Ser Leu Thr Thr Asp Pro Ser Gln Ser Leu Leu Ser  
 340 345 350

Leu Glu Gln Glu Pro Leu Leu Ser Arg Met Gly Ser Leu Arg Ala Pro  
 355 360 365

Val Asp Glu Asp Arg Leu Ser Pro Leu Val Ala Ala Asp Ser Leu Leu  
 370 375 380

Glu His Val Arg Glu Asp Phe Ser Gly Leu Leu Pro Glu Glu Phe Ile  
 385 390 395 400

Ser Leu Ser Pro Pro His Glu Ala Leu Asp Tyr His Phe Gly Leu Glu  
 405 410 415

Glu Gly Glu Gly Ile Arg Asp Leu Phe Asp Cys Asp Phe Gly Asp Leu  
 420 425 430

Thr Pro Leu Asp Phe  
 435

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2517 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GGAATTCCGT GGCCGGGACT TTGCAGGCAG CGGCGGCCGG GGGCGGAGCG GGATCGAGCC	60
CTCGCCGAGG CCTGCCGCCA TGGGCCCGCG CCGCCGCCGC CGCCTGTCAC CCGGGCCGCG	120
CGGGCCGTGA GCGTCATGGC CTTGGCCGGG GCCCCTGCGG GCGGCCCCATG CGCGCCGGCG	180
CTGGAGGCCC TGCTCGGGGC CGGCGCGCTG CGGCTGCTCG ACTCCTCGCA GATCGTCATC	240
ATCTCCGCCG CGCAGGACGC CAGCGCCCCG CCGGCTCCCA CCGGCCCCGC GGCGCCCGCC	300
GCCGGCCCCCT GCGACCCTGA CCTGCTGCTC TTCGCCACAC CGCAGGCGCC CCGGCCCCACA	360
CCCAGTGCGC CGCGGCCCGC GCTCGGCCGC CCGCCGGTGA AGCGGAGGCT GGACCTGGAA	420
ACTGACCATC AGTACCTGGC CGAGAGCAGT GGGCCAGCTC GGGGCAGAGG CCGCCATCCA	480
GGAAAAGGTG TGAAATCCCC GGGGGAGAAG TCACGCTATG AGACCTCACT GAATCTGACC	540
ACCAAGCGCT TCCTGGAGCT GCTGAGCCAC TCGGCTGACG GTGTCGTCGA CCTGAACTGG	600
GCTGCCGAGG TGCTGAAGGT GCAGAAGCGG CGCATCTATG ACATCACCAA CGTCCTTGAG	660
GGCATCCAGC TCATTGCCAA GAAGTCCAAG AACCACATCC AGTGGCTGGG CAGCCACACC	720
ACAGTGGGCG TCGGCGGACG GCTTGAGGGG TTGACCCAGG ACCTCCGACA GCTGCAGGAG	780
AGCGAGCAGC AGCTGGACCA CCTGATGAAT ATCTGTACTA CGCAGCTGCG CCTGCTCTCC	840

GAGGACACTG ACAGCCAGCG CCTGGCCTAC GTGACGTGTC AGGACCTTCG TAGCATTGCA 900  
 GACCCTGCAG AGCAGATGGT TATGGTGATC AAAGCCCCTC CTGAGACCCA GCTCCAAGCC 960  
 GTGGACTCTT CGGAGAACTT TCAGATCTCC CTTAAGAGCA AACAAGGCCC GATCGATGTT 1020  
 TTCCTGTGCC CTGAGGAGAC CGTAGGTGGG ATCAGCCCTG GGAAGACCCC ATCCCAGGAG 1080  
 GTCATTCTG AGGAGGAGAA CAGGGCCACT GACTCTGCCA CCATAGTGTC ACCACCACCA 1140  
 TCATCTCCCC CCTCATCCCT CACCACAGAT CCCAGCCAGT CTCTACTCAG CCTGGAGCAA 1200  
 GAACCGCTGT TGTCCCGGAT GGGCAGCCTG CGGGCTCCCG TGGACGAGGA CCGCCTGTCC 1260  
 CCGCTGGTGG CGGCCGACTC GCTCCTGGAG CATGTGCGGG AGGACTTCTC CGGCCTCCTC 1320  
 CCTGAGGAGT TCATCAGCCT TTCCCCACCC CACGAGGCCC TCGACTACCA CTTCGGCCTC 1380  
 GAGGAGGGCG AGGGCATCAG AGACCTCTTC GACTGTGACT TTGGGGACCT CACCCCCCTG 1440  
 GATTTCTGAC AGGGCTTGGA GGGACCAGGG TTTCCAGAGT AGCTCACCTT GTCTCTGCAG 1500  
 CCCTGGAGCC CCCTGTCCCT GGCCGTCCTC CCAGCCTGTT TGGAAACATT TAATTTATAC 1560  
 CCCTCTCCTC TGTCTCCAGA AGCTTCTAGC TCTGGGGTCT GGCTACCGCT AGGAGGCTGA 1620  
 GCAAGCCAGG AAGGGAAGGA GTCTGTGTGG TGTGTATGTG CATGCAGCCT ACACCCACAC 1680  
 GTGTGTACCG GGGGTGAATG TGTGTGAGCA TGTGTGTGTG CATGTACCGG GGAATGAAGG 1740  
 TGAACATACA CCTCTGTGTG TGCACTGCAG ACACGCCCCA GTGTGTCCAC ATGTGTGTGC 1800  
 ATGAGTCCAT CTCTGCGCGT GGGGGGGCTC TAACTGCACT TTCGGCCCTT TTGCTCGTGG 1860  
 GGTCCCACAA GGCCAGGGC AGTGCCTGCT CCCAGAATCT GGTGCTCTGA CCAGGCCAGG 1920  
 TGGGGAGGCT TTGGCTGGCT GGGCGTGTAG GACGGTGAGA GCACTTCTGT CTAAAGGTT 1980  
 TTTTCTGATT GAAGCTTTAA TGGAGCGTTA TTTATTTATC GAGGCCTCTT TGGTGAGCCT 2040  
 GGGGAATCAG CAAAAGGGGA GGAGGGGTGT GGGGTTGATA CCCCAACTCC CTCTACCTT 2100  
 GAGCAAGGGC AGGGGTCCCT GAGCTGTTCT TCTGCCCCAT ACTGAAGGAA CTGAGGCCTG 2160  
 GGTGATTTAT TTATTGGGAA AGTGAGGGAG GGAGACAGAC TGA CTGACAG CCATGGGTGG 2220  
 TCAGATGGTG GGGTGGGCCC TCTCCAGGGG GCCAGTTCAG GGCCAGCTG CCCCCAGGA 2280  
 TGGATATGAG ATGGGAGAGG TGAGTGGGGG ACCTTCACTG ATGTGGGCAG GAGGGGTGGT 2340  
 GAAGGCCTCC CCCAGCCCAG ACCCTGTGGT CCCTCCTGCA GTGTCTGAAG CGCCTGCCTC 2400  
 CCCACTGCTC TGCCCCACCC TCCAATCTGC ACTTTGATTT GCTTCCTAAC AGCTCTGTTC 2460  
 CCTCCTGCTT TGGTTTTAAT AAATATTTTG ATGACGTAA AAAAAGGAAT TCGATAT 2517

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2994 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TTCCGGTTTT	TCTCAGGGGA	CGTTGAAATT	ATTTTTGTAA	CGGGAGTCGG	GAGAGGACGG	60
GGCGTGCCCC	GCGTGCGCGC	GCGTCGTCCT	CCCCGGCGCT	CCTCCACAGC	TCGCTGGCTC	120
CCGCCGCGGA	AAGGCGTCAT	GCCGCCCAAA	ACCCCCCGAA	AAACGGCCGC	CACCGCCGCC	180
GCTGCCGCCG	CGGAACCCCC	GGCACCGCCG	CCGCCGCCCC	CTCCTGAGGA	GGACCCAGAG	240
CAGGACAGCG	GGCCGGAGGA	CCTGCCTCTC	GTCAGGCTTG	AGTTTGAAGA	AACAGAAGAA	300
CCTGATTTTA	CTGCATTATG	TCAGAAATTA	AAGATACCAG	ATCATGTCAG	AGAGAGAGCT	360
TGGTTAACTT	GGGAGAAAGT	TTCATCTGTG	GATGGAGTAT	TGGGAGGTTA	TATTCAAAAG	420
AAAAAGGAAC	TGTGGGGAAT	CTGTATCTTT	ATTGCAGCAG	TTGACCTAGA	TGAGATGTCG	480
TTCACTTTTA	CTGAGCTACA	GAAAAACATA	GAAATCAGTG	TCCATAAATT	CTTTAACTTA	540
CTAAAAGAAA	TTGATACCAG	TACCAAAGTT	GATAATGCTA	TGTCAAGACT	GTTGAAGAAG	600
TATGATGTAT	TGTTTGCACT	CTTCAGCAAA	TTGGAAAGGA	CATGTGAACT	TATATATTTG	660
ACACAACCCA	GCAGTTCGAT	ATCTACTGAA	ATAAATTCTG	CATTGGTGCT	AAAAGTTTCT	720
TGGATCACAT	TTTTATTAGC	TAAAGGGGAA	GTATTACAAA	TGGAAGATGA	TCTGGTGATT	780
TCATTTAGT	TAATGCTATG	TGTCCTTGAC	TATTTTATTA	AACTCTCACC	TCCCATGTTG	840
CTCAAAGAAC	CATATAAAAC	AGCTGTTATA	CCCATTAATG	GTTACCTCG	AACACCCAGG	900
CGAGGTCAGA	ACAGGAGTGC	ACGGATAGCA	AAACAAC TAG	AAAATGATAC	AAGAATTATT	960
GAAGTTCTCT	GTAAAGAACA	TGAATGTAAT	ATAGATGAGG	TGAAAAATGT	TTATTTCAAA	1020
AATTTTATAC	CTTTTATGAA	TTCTCTTGGA	CTTGTAACAT	CTAATGGACT	TCCAGAGGTT	1080
GAAAATCTTT	CTAAACGATA	CGAAGAAATT	TATCTTAAAA	ATAAAGATCT	AGATGCAAGA	1140
TTATTTTGG	ATCATGATAA	AACTCTTCAG	ACTGATTCTA	TAGACAGTTT	TGAAACACAG	1200
AGAACACCAC	GAAAAAGTAA	CCTTGATGAA	GAGGTGAATG	TAATTCCTCC	ACACACTCCA	1260
GTTAGGACTG	TTATGAACAC	TATCCAACAA	TTAATGATGA	TTTTAAATTC	AGCAAGTGAT	1320
CAACCTTCAG	AAAATCTGAT	TTCTATTTT	AACAAC TGCA	CAGTGAATCC	AAAAGAAAGT	1380
ATACTGAAAA	GAGTGAAGGA	TATAGGATAC	ATCTTTAAAG	AGAAATTTGC	TAAAGCTGTG	1440
GGACAGGGTT	GTGTCGAAAT	TGGATCACAG	CGATACAAAC	TTGGAGTTTCG	CTTGTATTAC	1500
CGAGTAATGG	AATCCATGCT	TAAATCAGAA	GAAGAACGAT	TATCCATTCA	AAATTTTAGC	1560
AAACTTCTGA	ATGACAACAT	TTTTCATATG	TCTTTATTGG	CGTGCGCTCT	TGAGGTTGTA	1620

ATGGCCACAT ATAGCAGAAG TACATCTCAG AATCTTGATT CTGGAACAGA TTTGTCTTTC 1680  
 CCATGGATTG TGAATGTGCT TAATTTAAAA GCCTTTGATT TTTACAAAGT GATCGAAAGT 1740  
 TTTATCAAAG CAGAAGGCAA CTTGACAAGA GAAATGATAA AACATTTAGA ACGATGTGAA 1800  
 CATCGAATCA TGGAATCCCT TGCATGGCTC TCAGATTCAC CTTTATTTGA TCTTATTTAA 1860  
 CAATCAAAGG ACCGAGAAGG ACCAACTGAT CACCTTGAAT CTGCTTGTCC TCTTAATCTT 1920  
 CCTCTCCAGA ATAATCACAC TGCAGCAGAT ATGTATCTTT CTCCTGTAAG ATCTCCAAAG 1980  
 AAAAAAGGTT CAACTACGCG TGTAATTCT ACTGCAAATG CAGAGACACA AGCAACCTCA 2040  
 GCCTTCCAGA CCCAGAAGCC ATTGAAATCT ACCTCTCTTT CACTGTTTTA TAAAAAAGTG 2100  
 TATCGGCTAG CCTATCTCCG GCTAAATACA CTTTGTGAAC GCCTTCTGTC TGAGCACCCA 2160  
 GAATTAGAAC ATATCATCTG GACCCTTTTT CAGCACACCC TGCAGAATGA GTATGAACTC 2220  
 ATGAGAGACA GGCATTTGGA CCAAATTATG ATGTGTTCCA TGTATGGCAT ATGCAAAGTG 2280  
 AAGAATATAG ACCTTAAATT CAAAATCATT GTAACAGCAT ACAAGGATCT TCCTCATGCT 2340  
 GTTCAGGAGA CATTCAAACG TGTTTTGATC AAAGAAGAGG AGTATGATTG TATTATAGTA 2400  
 TTCTATAACT CGGTCTTCAT GCAGAGACTG AAAACAAATA TTTTGCAGTA TGCTTCCACC 2460  
 AGGCCCCCTA CTTGTCCACC AATACCTCAC ATTCCTCGAA GCCCTTACAA GTTTCCTAGT 2520  
 TCACCCTTAC GGATTCCTGG AGGGAACATC TATATTTCAC CCCTGAAGAG TCCATATAAA 2580  
 ATTCAGAAAG GTCTGCCAAC ACCAACAAAA ATGACTCCAA GATCAAGAAT CTTAGTATCA 2640  
 ATTGGTGAAT CATTGGGAC TTCTGAGAAG TTCCAGAAAA TAAATCAGAT GGTATGTAAC 2700  
 AGCGACCGTG TGCTCAAAAG AAGTGCTGAA GGAAGCAACC CTCCTAAACC ACTGAAAAAA 2760  
 CTACGCTTTG ATATTGAAGG ATCAGATGAA GCAGATGGAA GTAAACATCT CCCAGGAGAG 2820  
 TCCAAATTTG AGCAGAAACT GGCAGAAATG ACTTCTACTC GAACACGAAT GCAAAGCAG 2880  
 AAAATGAATG ATAGCATGGA TACCTCAAAC AAGGAAGAGA AATGAGGATC TCAGGACCTT 2940  
 GGTGGACACT GTGTACACCT CTGGATTCAT TGTCTCTCAC AGATGTGACT GTAT 2994

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 928 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala Ala Ala

1	5				10				15					
Ala	Ala	Glu	Pro	Pro	Ala	Pro	Pro	Pro	Pro	Pro	Pro	Glu	Glu	Asp
			20					25				30		
Pro	Glu	Gln	Asp	Ser	Gly	Pro	Glu	Asp	Leu	Pro	Leu	Val	Arg	Leu
		35					40					45		Glu
Phe	Glu	Glu	Thr	Glu	Glu	Pro	Asp	Phe	Thr	Ala	Leu	Cys	Gln	Lys
	50					55					60			Leu
Lys	Ile	Pro	Asp	His	Val	Arg	Glu	Arg	Ala	Trp	Leu	Thr	Trp	Glu
65					70					75				80
Val	Ser	Ser	Val	Asp	Gly	Val	Leu	Gly	Gly	Tyr	Ile	Gln	Lys	Lys
				85					90					95
Glu	Leu	Trp	Gly	Ile	Cys	Ile	Phe	Ile	Ala	Ala	Val	Asp	Leu	Asp
			100					105					110	Glu
Met	Ser	Phe	Thr	Phe	Thr	Glu	Leu	Gln	Lys	Asn	Ile	Glu	Ile	Ser
		115					120					125		Val
His	Lys	Phe	Phe	Asn	Leu	Leu	Lys	Glu	Ile	Asp	Thr	Ser	Thr	Lys
	130					135					140			Val
Asp	Asn	Ala	Met	Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	Leu	Phe
145					150					155				160
Leu	Phe	Ser	Lys	Leu	Glu	Arg	Thr	Cys	Glu	Leu	Ile	Tyr	Leu	Thr
				165					170					175
Pro	Ser	Ser	Ser	Ile	Ser	Thr	Glu	Ile	Asn	Ser	Ala	Leu	Val	Leu
			180					185					190	Lys
Val	Ser	Trp	Ile	Thr	Phe	Leu	Leu	Ala	Lys	Gly	Glu	Val	Leu	Gln
		195					200					205		Met
Glu	Asp	Asp	Leu	Val	Ile	Ser	Phe	Gln	Leu	Met	Leu	Cys	Val	Leu
	210					215					220			Asp
Tyr	Phe	Ile	Lys	Leu	Ser	Pro	Pro	Met	Leu	Leu	Lys	Glu	Pro	Tyr
225					230					235				240
Thr	Ala	Val	Ile	Pro	Ile	Asn	Gly	Ser	Pro	Arg	Thr	Pro	Arg	Arg
				245					250					255
Gln	Asn	Arg	Ser	Ala	Arg	Ile	Ala	Lys	Gln	Leu	Glu	Asn	Asp	Thr
			260					265					270	Arg
Ile	Ile	Glu	Val	Leu	Cys	Lys	Glu	His	Glu	Cys	Asn	Ile	Asp	Glu
		275					280					285		Val
Lys	Asn	Val	Tyr	Phe	Lys	Asn	Phe	Ile	Pro	Phe	Met	Asn	Ser	Leu
	290					295					300			Gly
Leu	Val	Thr	Ser	Asn	Gly	Leu	Pro	Glu	Val	Glu	Asn	Leu	Ser	Lys
305					310					315				320
Tyr	Glu	Glu	Ile	Tyr	Leu	Lys	Asn	Lys	Asp	Leu	Asp	Ala	Arg	Leu
				325					330					335

Leu	Asp	His	Asp	Lys	Thr	Leu	Gln	Thr	Asp	Ser	Ile	Asp	Ser	Phe	Glu
			340						345						350
Thr	Gln	Arg	Thr	Pro	Arg	Lys	Ser	Asn	Leu	Asp	Glu	Glu	Val	Asn	Val
		355					360					365			
Ile	Pro	Pro	His	Thr	Pro	Val	Arg	Thr	Val	Met	Asn	Thr	Ile	Gln	Gln
	370					375					380				
Leu	Met	Met	Ile	Leu	Asn	Ser	Ala	Ser	Asp	Gln	Pro	Ser	Glu	Asn	Leu
385					390					395					400
Ile	Ser	Tyr	Phe	Asn	Asn	Cys	Thr	Val	Asn	Pro	Lys	Glu	Ser	Ile	Leu
				405					410					415	
Lys	Arg	Val	Lys	Asp	Ile	Gly	Tyr	Ile	Phe	Lys	Glu	Lys	Phe	Ala	Lys
			420					425					430		
Ala	Val	Gly	Gln	Gly	Cys	Val	Glu	Ile	Gly	Ser	Gln	Arg	Tyr	Lys	Leu
		435					440					445			
Gly	Val	Arg	Leu	Tyr	Tyr	Arg	Val	Met	Glu	Ser	Met	Leu	Lys	Ser	Glu
	450					455					460				
Glu	Glu	Arg	Leu	Ser	Ile	Gln	Asn	Phe	Ser	Lys	Leu	Leu	Asn	Asp	Asn
465					470					475					480
Ile	Phe	His	Met	Ser	Leu	Leu	Ala	Cys	Ala	Leu	Glu	Val	Val	Met	Ala
				485					490					495	
Thr	Tyr	Ser	Arg	Ser	Thr	Ser	Gln	Asn	Leu	Asp	Ser	Gly	Thr	Asp	Leu
			500					505					510		
Ser	Phe	Pro	Trp	Ile	Leu	Asn	Val	Leu	Asn	Leu	Lys	Ala	Phe	Asp	Phe
		515					520					525			
Tyr	Lys	Val	Ile	Glu	Ser	Phe	Ile	Lys	Ala	Glu	Gly	Asn	Leu	Thr	Arg
	530					535					540				
Glu	Met	Ile	Lys	His	Leu	Glu	Arg	Cys	Glu	His	Arg	Ile	Met	Glu	Ser
545					550					555					560
Leu	Ala	Trp	Leu	Ser	Asp	Ser	Pro	Leu	Phe	Asp	Leu	Ile	Lys	Gln	Ser
				565					570					575	
Lys	Asp	Arg	Glu	Gly	Pro	Thr	Asp	His	Leu	Glu	Ser	Ala	Cys	Pro	Leu
			580					585					590		
Asn	Leu	Pro	Leu	Gln	Asn	Asn	His	Thr	Ala	Ala	Asp	Met	Tyr	Leu	Ser
		595					600					605			
Pro	Val	Arg	Ser	Pro	Lys	Lys	Lys	Gly	Ser	Thr	Thr	Arg	Val	Asn	Ser
	610					615						620			
Thr	Ala	Asn	Ala	Glu	Thr	Gln	Ala	Thr	Ser	Ala	Phe	Gln	Thr	Gln	Lys
625					630					635					640
Pro	Leu	Lys	Ser	Thr	Ser	Leu	Ser	Leu	Phe	Tyr	Lys	Lys	Val	Tyr	Arg
				645					650					655	
Leu	Ala	Tyr	Leu	Arg	Leu	Asn	Thr	Leu	Cys	Glu	Arg	Leu	Leu	Ser	Glu
			660					665					670		



His Pro Glu Leu Glu His Ile Ile Trp Thr Leu Phe Gln His Thr Leu  
 675 680 685  
 Gln Asn Glu Tyr Glu Leu Met Arg Asp Arg His Leu Asp Gln Ile Met  
 690 695 700  
 Met Cys Ser Met Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp Leu Lys  
 705 710 715 720  
 Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala Val Gln  
 725 730 735  
 Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp Ser Ile  
 740 745 750  
 Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys Thr Asn Ile  
 755 760 765  
 Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile Pro His  
 770 775 780  
 Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile Pro  
 785 790 795 800  
 Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile Ser  
 805 810 815  
 Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg Ile Leu  
 820 825 830  
 Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys Ile  
 835 840 845  
 Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser Ala Glu  
 850 855 860  
 Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile Glu  
 865 870 875 880  
 Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser Lys  
 885 890 895  
 Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln  
 900 905 910  
 Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys  
 915 920 925

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3853 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 209..250

- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 254..289
- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 293..505
- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 509..514
- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 518..520
- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 524..658
- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 662..691
- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 695..748
- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 752..781
- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 785..829
- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 1132..1134
- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 1138..1149
- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 833..862

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GACGGATCGG GAGATCTCCC GATCCCCTAT GGTCGACTCT CAGTACAATC TGCTCTGATG	60
CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT GGAGGTCGCT GAGTAGTGCG	120
CGAGCAAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA CAATTGCATG AAGAATCTGC	180
TTAGGGTTAG GCGTTTTGCG CTGCTTCG CGA TGT ACG GGC CAG ATA TAC GCG	232
Arg Cys Thr Gly Gln Ile Tyr Ala	
1 5	
TTG ACA TTG ATT ATT GAC TAG TTA TTA ATA GTA ATC AAT TAC GGG GTC	280

Leu	Thr	Leu	Ile	Ile	Asp		Leu	Leu	Ile	Val	Ile	Asn	Tyr	Gly	Val		
10							1				5						
ATT	AGT	TCA	TAG	CCC	ATA	TAT	GGA	GTT	CCG	CGT	TAC	ATA	ACT	TAC	GGT	328	
Ile	Ser	Ser		Pro	Ile	Tyr	Gly	Val	Pro	Arg	Tyr	Ile	Thr	Tyr	Gly		
10				1				5					10				
AAA	TGG	CCC	GCC	TGG	CTG	ACC	GCC	CAA	CGA	CCC	CCG	CCC	ATT	GAC	GTC	376	
Lys	Trp	Pro	Ala	Trp	Leu	Thr	Ala	Gln	Arg	Pro	Pro	Pro	Ile	Asp	Val		
		15					20					25					
AAT	AAT	GAC	GTA	TGT	TCC	CAT	AGT	AAC	GCC	AAT	AGG	GAC	TTT	CCA	TTG	424	
Asn	Asn	Asp	Val	Cys	Ser	His	Ser	Asn	Ala	Asn	Arg	Asp	Phe	Pro	Leu		
	30					35					40						
ACG	TCA	ATG	GGT	GGA	CTA	TTT	ACG	GTA	AAC	TGC	CCA	CTT	GGC	AGT	ACA	472	
Thr	Ser	Met	Gly	Gly	Leu	Phe	Thr	Val	Asn	Cys	Pro	Leu	Gly	Ser	Thr		
45					50					55					60		
TCA	AGT	GTA	TCA	TAT	GCC	AAG	TAC	GCC	CCC	TAT	TGA	CGT	CAA			514	
Ser	Ser	Val	Ser	Tyr	Ala	Lys	Tyr	Ala	Pro	Tyr		Arg	Gln				
				65					70			1					
TGA	CGG	TAA	ATG	GCC	CGC	CTG	GCA	TTA	TGC	CCA	GTA	CAT	GAC	CTT	ATG	562	
	Arg		Met	Ala	Arg	Leu	Ala	Leu	Cys	Pro	Val	His	Asp	Leu	Met		
	1		1				5					10					
GGA	CTT	TCC	TAC	TTG	GCA	GTA	CAT	CTA	CGT	ATT	AGT	CAT	CGC	TAT	TAC	610	
Gly	Leu	Ser	Tyr	Leu	Ala	Val	His	Leu	Arg	Ile	Ser	His	Arg	Tyr	Tyr		
	15					20					25						
CAT	GGT	GAT	GCG	GTT	TTG	GCA	GTA	CAT	CAA	TGG	GCG	TGG	ATA	GCG	GTT	658	
His	Gly	Asp	Ala	Val	Leu	Ala	Val	His	Gln	Trp	Ala	Trp	Ile	Ala	Val		
30					35					40					45		
TGA	CTC	ACG	GGG	ATT	TCC	AAG	TCT	CCA	CCC	CAT	TGA	CGT	CAA	TGG	GAG	706	
	Leu	Thr	Gly	Ile	Ser	Lys	Ser	Pro	Pro	His		Arg	Gln	Trp	Glu		
	1				5					10		1					
TTT	GTT	TTG	GCA	CCA	AAA	TCA	ACG	GGA	CTT	TCC	AAA	ATG	TCG			748	
Phe	Val	Leu	Ala	Pro	Lys	Ser	Thr	Gly	Leu	Ser	Lys	Met	Ser				
5					10					15							
TAA	CAA	CTC	CGC	CCC	ATT	GAC	GCA	AAT	GGG	CGG	TAG	CGC	TGT	ACG	GTG	796	
	Gln	Leu	Arg	Pro	Ile	Asp	Ala	Asn	Gly	Arg		Arg	Cys	Thr	Val		
	1				5					10		1					
GGA	GGT	CTA	TAT	AAG	CAG	AGC	TCT	CTG	GCT	AAC	TAG	AGA	ACC	CAC	TGC	844	
Gly	Gly	Leu	Tyr	Lys	Gln	Ser	Ser	Leu	Ala	Asn		Arg	Thr	His	Cys		
5					10					15		1					

CTA	TAG	TGT	CAC	CTA	AAT	GCTAGAGCTC	GCTGATCAGC	CTCGACTGTG	
Leu		Cys	His	Leu	Asn				1179
1		1							
CCTTCTAGTT	GCCAGCCATC	TGTTGTTTGC	CCCTCCCCCG	TGCCTTCCTT	GACCCTGGAA				1239
GGTGCCACTC	CCACTGTCCT	TTCCTAATAA	AATGAGGAAA	TTGCATCGCA	TTGTCTGAGT				1299
AGGTGTCATT	CTATTCTGGG	GGGTGGGGTG	GGGCAGGACA	GCAAGGGGGA	GGATTGGGAA				1359
GACAATAGCC	GAAATGACCG	ACCAAGCGAC	GCCCAACCTG	CCATCACGAG	ATTTTCGATTC				1419
CACCGCCGCC	TTCTATGAAA	GGTTGGGCTT	CGGAATCGTT	TTCCGGGACG	CCGGCTGGAT				1479
GATCCTCCAG	CGCGGGGATC	TCATGCTGGA	GTTCTTCGCC	CACCCCAACT	TGTTTATTGC				1539
AGCTTATAAT	GGTTACAAAT	AAAGCAATAG	CATCACAAAT	TTCACAAATA	AAGCATTTTT				1599
TTCCTGTCAT	TCTAGTTGTG	GTTTGTCCAA	ACTCATCAAT	GATCTTATC	ATGTCTGTAT				1659
ACCGTCGACC	TCTAGCTAGA	GCTTGGCGTA	ATCATGGTCA	TAGCTGTTTC	CTGTGTGAAA				1719
TTGTTATCCG	CTCACAATTC	CACACAACAT	ACGAGCCGGA	AGCATAAAGT	GTAAAGCCTG				1779
GGGTGCCTAA	TGAGTGAGCT	AACTCACATT	AATTGCGTTG	CGCTCACTGC	CCGCTTTCCA				1839
GTCGGGAAAC	CTGTCGTGCC	AGCTGCATTA	ATGAATCGGC	CAACGCGCGG	GGAGAGGCGG				1899
TTTGCGTATT	GGGCGCTCTT	CCGCTTCCTC	GCTCACTGAC	TCGCTGCGCT	CGGTGCTTCG				1959
GCTGCGGCGA	GCGGTATCAG	CTCACTCAAA	GGCGGTAATA	CGGTTATCCA	CAGAATCAGG				2019
GGATAACGCA	GGAAAGAACA	TGTGAGCAAA	AGGCCAGCAA	AAGGCCAGGA	ACCGTAAAAA				2079
GGCCGCGTTG	CTGGCGTTTT	TCCATAGGCT	CCGCCCCCCT	GACGAGCATC	ACAAAAATCG				2139
ACGCTCAAGT	CAGAGGTGGC	GAAACCCGAC	AGGACTATAA	AGATACCAGG	CGTTTCCCCC				2199
TGGAAGCTCC	CTCGTGCGCT	CTCCTGTTCC	GACCCTGCCG	CTTACCGGAT	ACCTGTCCGC				2259
CTTTCTCCCT	TCGGGAAGCG	TGGCGCTTTC	TCAATGCTCA	CGCTGTAGGT	ATCTCAGTTC				2319
GGTGTAGGTC	GTTCGCTCCA	AGCTGGGCTG	TGTGCACGAA	CCCCCGTTC	AGCCCGACCG				2379
CTGCGCCTTA	TCCGGTAACT	ATCGTCTTGA	GTCCAACCCG	GTAAGACACG	ACTTATCGCC				2439
ACTGGCAGCA	GCCACTGGTA	ACAGGATTAG	CAGAGCGAGG	TATGTAGGCG	GTGCTACAGA				2499
GTTCTTGAAG	TGGTGGCCTA	ACTACGGCTA	CACTAGAAGG	ACAGTATTTG	GTATCTGCGC				2559
TCTGCTGAAG	CCAGTTACCT	TCGGAAAAAG	AGTTGGTAGC	TCTTGATCCG	GCAAACAAAC				2619
CACCGCTGGT	AGCGGTGGTT	TTTTTGTTTG	CAAGCAGCAG	ATTACGCGCA	GAAAAAAAGG				2679
ATCTCAAGAA	GATCCTTTGA	TCTTTTCTAC	GGGGTCTGAC	GCTCAGTGGA	ACGAAAACTC				2739
ACGTTAAGGG	ATTTTGGTCA	TGAGATTATC	AAAAAGGATC	TTCACCTAGA	TCCTTTTAAA				2799
TTAAAAATGA	AGTTTTTAAAT	CAATCTAAAG	TATATATGAG	TAAACTTGGT	CTGACAGTTA				2859
CCAATGCTTA	ATCAGTGAGG	CACCTATCTC	AGCGATCTGT	CTATTTTCGTT	CATCCATAGT				2919

TGCCTGACTC CCCGTCGTGT AGATAACTAC GATACGGGAG GGCTTACCAT CTGGCCCCAG 2979  
 TGCTGCAATG ATACCGCGAG ACCCACGCTC ACCGGCTCCA GATTTATCAG CAATAAACCA 3039  
 GCCAGCCGGA AGGGCCGAGC GCAGAAGTGG TCCTGCAACT TTATCCGCCT CCATCCAGTC 3099  
 TATTAATTGT TGCCGGGAAG CTAGAGTAAG TAGTTCGCCA GTTAATAGTT TGCGCAACGT 3159  
 TGTTGCCATT GCTACAGGCA TCGTGGTGTC ACGCTCGTCG TTTGGTATGG CTTCATTCAG 3219  
 CTCCGGTTCC CAACGATCAA GGCGAGTTAC ATGATCCCCC ATGTTGTGCA AAAAAGCGGT 3279  
 TAGCTCCTTC GGTCTCCGA TCGTTGTCAG AAGTAAGTTG GCCGCAGTGT TATCACTCAT 3339  
 GGTTATGGCA GCACTGCATA ATTCTCTTAC TGTCATGCCA TCCGTAAGAT GCTTTTCTGT 3399  
 GACTGGTGAG TACTCAACCA AGTCATTCTG AGAATAGTGT ATGCGGCGAC CGAGTTGCTC 3459  
 TTGCCCGGCG TCAATACGGG ATAATACCGC GCCACATAGC AGAACTTTAA AAGTGCTCAT 3519  
 CATTGGAAAA CGTCTTCGG GGCAGAACT CTCAAGGATC TTACCGCTGT TGAGATCCAG 3579  
 TTCGATGTAA CCCACTCGTG CACCCAACCTG ATCTTCAGCA TCTTTTACTT TCACCAGCGT 3639  
 TTCTGGGTGA GCAAAAACAG GAAGGCAAAA TGCCGCAAAA AAGGGAATAA GGGCGACACG 3699  
 GAAATGTTGA ATACTCATAC TCTTCCTTTT TCAATATTAT TGAAGCATTT ATCAGGGTTA 3759  
 TTGTCTCATG AGCGGATACA TATTTGAATG TATTTAGAAA AATAAACAAA TAGGGGTTCC 3819  
 GCGCACATTT CCCCAGAAAAG TGCCACCTGA CGTC 3853

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Arg Cys Thr Gly Gln Ile Tyr Ala Leu Thr Leu Ile Ile Asp  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Leu Leu Ile Val Ile Asn Tyr Gly Val Ile Ser Ser  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 71 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```

Pro Ile Tyr Gly Val Pro Arg Tyr Ile Thr Tyr Gly Lys Trp Pro Ala
 1             5             10             15
Trp Leu Thr Ala Gln Arg Pro Pro Pro Ile Asp Val Asn Asn Asp Val
                20             25             30
Cys Ser His Ser Asn Ala Asn Arg Asp Phe Pro Leu Thr Ser Met Gly
                35             40             45
Gly Leu Phe Thr Val Asn Cys Pro Leu Gly Ser Thr Ser Ser Val Ser
 50             55             60
Tyr Ala Lys Tyr Ala Pro Tyr
 65             70

```

## (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```

Arg Gln
 1

```

## (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```

Arg
 1

```

## (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 45 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
Met Ala Arg Leu Ala Leu Cys Pro Val His Asp Leu Met Gly Leu Ser
 1           5           10           15
Tyr Leu Ala Val His Leu Arg Ile Ser His Arg Tyr Tyr His Gly Asp
          20           25           30
Ala Val Leu Ala Val His Gln Trp Ala Trp Ile Ala Val
      35           40           45
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Leu Thr Gly Ile Ser Lys Ser Pro Pro His
 1           5           10
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
Arg Gln Trp Glu Phe Val Leu Ala Pro Lys Ser Thr Gly Leu Ser Lys
 1           5           10           15
```

Met Ser

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Gln Leu Arg Pro Ile Asp Ala Asn Gly Arg
 1           5           10
```

## (2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Arg	Cys	Thr	Val	Gly	Gly	Leu	Tyr	Lys	Gln	Ser	Ser	Leu	Ala	Asn
1				5					10					15

## (2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Arg	Thr	His	Cys	Leu	Leu	Ala	Tyr	Arg	Asn
1				5					10

## (2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Leu
1

## (2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 4 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Cys	His	Leu	Asn
1			

## (2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:



- (A) LENGTH: 4026 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 209..250

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 254..289

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 293..505

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 509..514

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 518..520

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 524..658

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 662..691

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 695..748

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 752..781

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 785..829

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 833..862

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1305..1307

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1311..1322

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted January 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

GACGGATCGG	GAGATCTCCC	GATCCCCTAT	GGTCGACTCT	CAGTACAATC	TGCTCTGATG		60
CCGCATAGTT	AAGCCAGTAT	CTGCTCCCTG	CTTGTGTGTT	GGAGGTCGCT	GAGTAGTGCG		120
CGAGCAAAAT	TTAAGCTACA	ACAAGGCAAG	GCTTGACCGA	CAATTGCATG	AAGAATCTGC		180
TTAGGGTTAG	GCGTTTTGCG	CTGCTTCG	CGA	TGT	ACG	GGC	232
			Arg	Cys	Thr	Gly	
			1			5	
TTG	ACA	TTG	ATT	ATT	GAC	TAG	280
Leu	Thr	Leu	Ile	Ile	Asp	Leu	
	10					1	
ATT	AGT	TCA	TAG	CCC	ATA	TAT	328
Ile	Ser	Ser		Pro	Ile	Tyr	
	10			1		5	
AAA	TGG	CCC	GCC	TGG	CTG	ACC	376
Lys	Trp	Pro	Ala	Trp	Leu	Thr	
		15				20	
AAT	AAT	GAC	GTA	TGT	TCC	CAT	424
Asn	Asn	Asp	Val	Cys	Ser	His	
	30					35	
ACG	TCA	ATG	GGT	GGA	CTA	TTT	472
Thr	Ser	Met	Gly	Gly	Leu	Phe	
	45				50		
TCA	AGT	GTA	TCA	TAT	GCC	AAG	514
Ser	Ser	Val	Ser	Tyr	Ala	Lys	
				65		70	
TGA	CGG	TAA	ATG	GCC	CGC	CTG	562
	Arg		Met	Ala	Arg	Leu	
	1		1			5	
GGA	CTT	TCC	TAC	TTG	GCA	GTA	610
Gly	Leu	Ser	Tyr	Leu	Ala	Val	
	15					20	
CAT	GGT	GAT	GCG	GTT	TTG	GCA	658
His	Gly	Asp	Ala	Val	Leu	Ala	
	30				35		
TGA	CTC	ACG	GGG	ATT	TCC	AAG	706
	Leu	Thr	Gly	Ile	Ser	Lys	
	1				5		
TTT	GTT	TTG	GCA	CCA	AAA	TCA	748
Phe	Val	Leu	Ala	Pro	Lys	Ser	
	5				10		
TAA	CAA	CTC	CGC	CCC	ATT	GAC	796
	Gln	Leu	Arg	Pro	Ile	Asp	
	1				5		
GGA	GGT	CTA	TAT	AAG	CAG	AGC	844
Gly	Gly	Leu	Tyr	Lys	Gln	Ser	
	5				10		

TTA CTG GCT TAT CGA AAT TAATACGACT CACTATAGGG AGACCCAAGC	892
Leu Leu Ala Tyr Arg Asn	
5 10	
TTCGCGCGGG TACCACTCTC TTCCGCATCG CTGTCTGCGA GGGCCAGCTG TTGGGCTCGC	952
GGTTGAGGAC AAACCTCTTCG CGGTCTTTTC AGTACTCTTG GATCGGAAAC CCGTCGGCCT	1012
CCGAACGGTA CTCCGCCACC GAGGGACCTG AGCGAGTCCG CATCGACCGG ATCGGAAAAC	1072
CTCTCGAGGA ACTGAAAAAC CAGAAAGTTA ACTGGTAAGT TTAGTCTTTT TGTCTTTTAA	1132
TTTCAGGTCC CGGATCCGGT GGTGGTGCAA ATCAAAGAAC TGCTCCTCAG TGGATGTTGC	1192
CTTTACTTCT AGGCCTGTAC GGAAGTGTTA CTTCTGCTCT AAAAGCTGCG GAATTGTACC	1252
CGCGGCCGCT GCAGTCTAGA CGAATTCGCG TACGATATCG ATGGGCCCTA TT CTA	1307
Leu	
1	
TAG TGT CAC CTA AAT GCTAGAGCTC GCTGATCAGC CTCGACTGTG CCTTCTAGTT	1362
Cys His Leu Asn	
1	
GCCAGCCATC TGTGTTTTCG CCCTCCCCCG TGCCTTCCTT GACCCTGGAA GGTGCCACTC	1422
CCACTGTCCT TTCCTAATAA AATGAGGAAA TTGCATCGCA TTGTCTGAGT AGGTGTCATT	1482
CTATTCTGGG GGGTGGGGTG GGGCAGGACA GCAAGGGGGA GGATTGGGAA GACAATAGCC	1542
GAAATGACCG ACCAAGCGAC GCCCAACCTG CCATCACGAG ATTTTCGATT CACCGCCGCC	1602
TTCTATGAAA GGTGGGGCTT CGGAATCGTT TTCCGGGACG CCGGCTGGAT GATCCTCCAG	1662
CGCGGGGATC TCATGCTGGA GTTCTTCGCC CACCCCAACT TGTTTATTGC AGCTTATAAT	1722
GGTTACAAAT AAAGCAATAG CATCACAAAT TTCACAAATA AAGCATTTTT TTTACTGCAT	1782
TCTAGTTGTG GTTTGTCCAA ACTCATCAAT GTATCTTATC ATGTCTGTAT ACCGTCGACC	1842
TCTAGCTAGA GCTTGGCGTA ATCATGGTCA TAGCTGTTTC CTGTGTGAAA TTGTTATCCG	1902
CTCACAATTC CACACAACAT ACGAGCCGGA AGCATAAAGT GTAAAGCCTG GGGTGCCTAA	1962
TGAGTGAGCT AACTCACATT AATTGCGTTG CGCTCACTGC CCGCTTTCCA GTCGGGAAAC	2022
CTGTGCTGCC AGCTGCATTA ATGAATCGGC CAACGCGCGG GGAGAGGCGG TTTGCGTATT	2082
GGGCGCTCTT CCGCTTCCTC GCTCACTGAC TCGCTGCGCT CGGTGCTTCG GCTGCGGCGA	2142
GCGGTATCAG CTCACTCAAA GGCGGTAATA CGGTTATCCA CAGAATCAGG GGATAACGCA	2202
GGAAAGAACA TGTGAGCAA AGGCCAGCAA AAGGCCAGGA ACCGTAAAAA GGCCGCGTTG	2262
CTGGCGTTTT TCCATAGGCT CCGCCCCCTT GACGAGCATC ACAAAAATCG ACGCTCAAGT	2322
CAGAGGTGGC GAAACCCGAC AGGACTATAA AGATACCAGG CGTTTCCCCC TGGAAGCTCC	2382
CTCGTGCGCT CTCCTGTTCC GACCCTGCCG CTTACCGGAT ACCTGTCCGC CTTTCTCCCT	2442
TCGGGAAGCG TGGCGCTTTC TCAATGCTCA CGCTGTAGGT ATCTGAGTTC GGTGTAGGTC	2502

GTTCGCTCCA AGCTGGGCTG TGTGCACGAA CCCCCCGTTC AGCCCGACCG CTGCGCCTTA 2562  
 TCCGGTAACT ATCGTCTTGA GTCCAACCCG GTAAGACACG ACTTATCGCC ACTGGCAGCA 2622  
 GCCACTGGTA ACAGGATTAG CAGAGCGAGG TATGTAGGCG GTGCTACAGA GTTCTTGAAG 2682  
 TGGTGGCCTA ACTACGGCTA CACTAGAAGG ACAGTATTTG GTATCTGCGC TCTGCTGAAG 2742  
 CCAGTTACCT TCGGAAAAAG AGTTGGTAGC TCTTGATCCG GCAAACAAAC CACCGCTGGT 2802  
 AGCGGTGGTT TTTTGTGTTG CAAGCAGCAG ATTACGCGCA GAAAAAAGG ATCTCAAGAA 2862  
 GATCCTTTGA TCTTTTCTAC GGGGTCTGAC GCTCAGTGGA ACGAAAACTC ACGTTAAGGG 2922  
 ATTTTGGTCA TGAGATTATC AAAAAGGATC TTCACCTAGA TCCTTTTAAA TTAAAAATGA 2982  
 AGTTTAAAT CAATCTAAAG TATATATGAG TAAACTTGGT CTGACAGTTA CCAATGCTTA 3042  
 ATCAGTGAGG CACCTATCTC AGCGATCTGT CTATTTCTGT CATCCATAGT TGCCTGACTC 3102  
 CCCGTCGTGT AGATAACTAC GATACGGGAG GGCTTACCAT CTGGCCCCAG TGCTGCAATG 3162  
 ATACCGCGAG ACCCAGCTC ACCGGCTCCA GATTTATCAG CAATAAACCA GCCAGCCGGA 3222  
 AGGGCCGAGC GCAGAAGTGG TCCTGCAACT TTATCCGCCT CCATCCAGTC TATTAATTGT 3282  
 TGCCGGGAAG CTAGAGTAAG TAGTTCGCCA GTTAATAGTT TGCGCAACGT TGTGCCATT 3342  
 GCTACAGGCA TCGTGGTGTC ACGCTCGTCG TTTGGTATGG CTTCAATCAG CTCCGGTTCC 3402  
 CAACGATCAA GGCGAGTTAC ATGATCCCCC ATGTTGTGCA AAAAAGCGGT TAGCTCCTTC 3462  
 GGTCTCCGA TCGTTGTCAG AAGTAAGTTG GCCGCAGTGT TATCACTCAT GGTTATGGCA 3522  
 GCACTGCATA ATTCTCTTAC TGTCATGCCA TCCGTAAGAT GCTTTTCTGT GACTGGTGAG 3582  
 TACTCAACCA AGTCATTCTG AGAATAGTGT ATGCGGCGAC CGAGTTGCTC TTGCCCCGCG 3642  
 TCAATACGGG ATAATACCGC GCCACATAGC AGAACTTTAA AAGTGCTCAT CATTGGAAAA 3702  
 CGTTCTTCGG GGCGAAAACT CTCAAGGATC TTACCGCTGT TGAGATCCAG TTCGATGTAA 3762  
 CCCACTCGTG CACCCAACCTG ATCTTCAGCA TCTTTTACTT TCACCAGCGT TTCTGGGTGA 3822  
 GCAAAAACAG GAAGGCAAAA TGCCGCAAAA AAGGGAATAA GGGCGACACG GAAATGTTGA 3882  
 ATACTCATAC TCTTCCTTTT TCAATATTAT TGAAGCATTT ATCAGGGTTA TTGTCTCATG 3942  
 AGCGGATACA TATTTGAATG TATTTAGAAA AATAAACAAA TAGGGGTTCC GCGCACATTT 4002  
 CCCCAGAAAAG TGCCACCTGA CGTC 4026

## (2) INFORMATION FOR SEQ ID NO:20:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Arg Cys Thr Gly Gln Ile Tyr Ala Leu Thr Leu Ile Ile Asp  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:21:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Leu Leu Ile Val Ile Asn Tyr Gly Val Ile Ser Ser  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:22:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 71 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Pro Ile Tyr Gly Val Pro Arg Tyr Ile Thr Tyr Gly Lys Trp Pro Ala  
 1 5 10 15  
 Trp Leu Thr Ala Gln Arg Pro Pro Pro Ile Asp Val Asn Asn Asp Val  
 20 25 30  
 Cys Ser His Ser Asn Ala Asn Arg Asp Phe Pro Leu Thr Ser Met Gly  
 35 40 45  
 Gly Leu Phe Thr Val Asn Cys Pro Leu Gly Ser Thr Ser Ser Val Ser  
 50 55 60  
 Tyr Ala Lys Tyr Ala Pro Tyr  
 65 70

## (2) INFORMATION FOR SEQ ID NO:23:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Arg Gln  
 1

## (2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Arg  
1

## (2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 45 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Met	Ala	Arg	Leu	Ala	Leu	Cys	Pro	Val	His	Asp	Leu	Met	Gly	Leu	Ser
1				5					10					15	
Tyr	Leu	Ala	Val	His	Leu	Arg	Ile	Ser	His	Arg	Tyr	Tyr	His	Gly	Asp
			20					25					30		
Ala	Val	Leu	Ala	Val	His	Gln	Trp	Ala	Trp	Ile	Ala	Val			
			35				40					45			

## (2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Leu	Thr	Gly	Ile	Ser	Lys	Ser	Pro	Pro	His
1				5					10

## (2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Arg Gln Trp Glu Phe Val Leu Ala Pro Lys Ser Thr Gly Leu Ser Lys  
 1 5 10 15

Met Ser

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Gln Leu Arg Pro Ile Asp Ala Asn Gly Arg  
 1 5 10

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Ala Cys Thr Val Gly Gly Leu Tyr Lys Gln Ser Ser Leu Ala Asn  
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Arg Thr His Cys Leu Leu Ala Tyr Arg Asn  
 1 5 10

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Leu  
1

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 4 amino acids  
    (B) TYPE: amino acid  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Cys His Leu Asn  
1

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 4249 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (ix) FEATURE:  
    (A) NAME/KEY: CDS  
    (B) LOCATION: 209..250

- (ix) FEATURE:  
    (A) NAME/KEY: CDS  
    (B) LOCATION: 254..289

- (ix) FEATURE:  
    (A) NAME/KEY: CDS  
    (B) LOCATION: 293..505

- (ix) FEATURE:  
    (A) NAME/KEY: CDS  
    (B) LOCATION: 509..514

- (ix) FEATURE:  
    (A) NAME/KEY: CDS  
    (B) LOCATION: 518..520

- (ix) FEATURE:  
    (A) NAME/KEY: CDS  
    (B) LOCATION: 524..658

- (ix) FEATURE:  
    (A) NAME/KEY: CDS  
    (B) LOCATION: 662..691

- (ix) FEATURE:  
    (A) NAME/KEY: CDS  
    (B) LOCATION: 695..748

(ix) FEATURE:

CC-0. In Public Domain. Not for Redistribution



(A) NAME/KEY: CDS  
(B) LOCATION: 752..781

## (ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 785..829

## (ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 833..862

## (ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 1528..1530

## (ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 1534..1545

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GACGGATCGG GAGATCTCCC GATCCCCTAT GGTCTGACTCT CAGTACAATC TGCTCTGATG	60
CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT GGAGGTCGCT GAGTAGTGCG	120
CGAGCAAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA CAATTGCATG AAGAATCTGC	180
TTAGGGTTAG GCGTTTTGCG CTGCTTCG CGA TGT ACG GGC CAG ATA TAC GCG	232
Arg Cys Thr Gly Gln Ile Tyr Ala	
1 5	
TTG ACA TTG ATT ATT GAC TAG TTA TTA ATA GTA ATC AAT TAC GGG GTC	280
Leu Thr Leu Ile Ile Asp Leu Leu Ile Val Ile Asn Tyr Gly Val	
10 1 5	
ATT AGT TCA TAG CCC ATA TAT GGA GTT CCG CGT TAC ATA ACT TAC GGT	328
Ile Ser Ser Pro Ile Tyr Gly Val Pro Arg Tyr Ile Thr Tyr Gly	
10 1 5 10	
AAA TGG CCC GCC TGG CTG ACC GCC CAA CGA CCC CCG CCC ATT GAC GTC	376
Lys Trp Pro Ala Trp Leu Thr Ala Gln Arg Pro Pro Ile Asp Val	
15 20 25	
AAT AAT GAC GTA TGT TCC CAT AGT AAC GCC AAT AGG GAC TTT CCA TTG	424
Asn Asn Asp Val Cys Ser His Ser Asn Ala Asn Arg Asp Phe Pro Leu	
30 35 40	
ACG TCA ATG GGT GGA CTA TTT ACG GTA AAC TGC CCA CTT GGC AGT ACA	472
Thr Ser Met Gly Gly Leu Phe Thr Val Asn Cys Pro Leu Gly Ser Thr	
45 50 55 60	
TCA AGT GTA TCA TAT GCC AAG TAC GCC CCC TAT TGA CGT CAA	514
Ser Ser Val Ser Tyr Ala Lys Tyr Ala Pro Tyr Arg Gln	
65 70 1	
TGA CGG TAA ATG GCC CGC CTG GCA TTA TGC CCA GTA CAT GAC CTT ATG	562
Arg Met Ala Arg Leu Ala Leu Cys Pro Val His Asp Leu Met	
1 1 5 10	
GGA CTT TCC TAC TTG GCA GTA CAT CTA CGT ATT AGT CAT CGC TAT TAC	610
Gly Leu Ser Tyr Leu Ala Val His Leu Arg Ile Ser His Arg Tyr Tyr	

15	20	25	
CAT GGT GAT GCG GTT TTG GCA GTA CAT CAA TGG GCG TGG ATA GCG GTT His Gly Asp Ala Val Leu Ala Val His Gln Trp Ala Trp Ile Ala Val 30 35 40 45			658
TGA CTC ACG GGG ATT TCC AAG TCT CCA CCC CAT TGA CGT CAA TGG GAG Leu Thr Gly Ile Ser Lys Ser Pro Pro His Arg Gln Trp Glu 1 5 10 1			706
TTT GTT TTG GCA CCA AAA TCA ACG GGA CTT TCC AAA ATG TCG Phe Val Leu Ala Pro Lys Ser Thr Gly Leu Ser Lys Met Ser 5 10 15			748
TAA CAA CTC CGC CCC ATT GAC GCA AAT GGG CGG TAG GCG TGT ACG GTG Gln Leu Arg Pro Ile Asp Ala Asn Gly Arg Ala Cys Thr Val 1 5 10 1			796
GGA GGT CTA TAT AAG CAG AGC TCT CTG GCT AAC TAG AGA ACC CAC TGC Gly Gly Leu Tyr Lys Gln Ser Ser Leu Ala Asn Arg Thr His Cys 5 10 15 1			844
TTA CTG GCT TAT CGA AAT TAATACGACT CACTATAGGG AGACCCAAGC Leu Leu Ala Tyr Arg Asn 5 10			892
TTTCGCGCGGG TACCACTCTC TTCCGCATCG CTGTCTGCGA GGGCCAGCTG TTGGGCTCGC			952
GGTTGAGGAC AAACCTCTTCG CGGTCTTTCC AGTACTCTTG GATCGGAAAC CCGTCGGCCT			1012
CCGAACGGTA CTCCGCCACC GAGGGACCTG AGCGAGTCCG CATCGACCGG ATCGGAAAAC			1072
CTCTCGAGGA ACTGAAAAAC CAGAAAGTTA ACTGGTAAGT TTAGTCTTTT TGTCTTTTTA			1132
TTTCAGGTCC CGGATCTGAG TTAGGGCGGG ACATGGGCGG AGTTAGGGGC GGGACTATGG			1192
TTGCTGACTA ATTGAGATGC ATGCTTTGCA TACTTCTGCC TGCTGGGGAG CCTGGGGACT			1252
TTCCACACCT GGTTGCTGAC TAATTGAGAT GCATGCTTTG CATACTTCTG CCTGCTGGGG			1312
AGCCTGGGGA CTTTCCACAC CCTAACTGAC ACACATTCCA CAGCTGGTTC TTTCAGATCC			1372
GGTGGTGGTG CAAATCAAAG AACTGCTCCT CAGTGGATGT TGCCTTTACT TCTAGGCCTG			1432
TACGGAAGTG TTAATTCTGC TCTAAAAGCT GCGGAATTGT ACCCGCGGCC GCTGCAGTCT			1492
AGACGAATTC GCGTACGATA TCGATGGGCC CTATT CTA TAG TGT CAC CTA AAT Leu Cys His Leu Asn 1 1			1545
GCTAGAGCTC GCTGATCAGC CTCGACTGTG CCTTCTAGTT GCCAGCCATC TGTTGTTTGC			1605
CCCTCCCCCG TGCCTTCCTT GACCCTGGAA GGTGCCACTC CCACTGTCCT TTCCTAATAA			1665
AATGAGGAAA TTGCATCGCA TTGTCTGAGT AGGTGTCATT CTATTCTGGG GGGTGGGGTG			1725
GGGCAGGACA GCAAGGGGGA GGATTGGGAA GACAATAGCC GAAATGACCG ACCAAGCGAC			1785
GCCCAACCTG CCATCACGAG ATTTTCGATTC CACCGCCGCC TTCTATGAAA GGTGGGGCTT			1845
CGGAATCGTT TTCCGGGACG CCGGCTGGAT GATCCTCCAG CGCGGGGATC TCATGCTGGA			1905

GTTCTTCGCC	CACCCCAACT	TGTTTATTGC	AGCTTATAAT	GGTTACAAAT	AAAGCAATAG	1965
CATCACAAAT	TTCACAAATA	AAGCATTTTT	TTCACTGCAT	TCTAGTTGTG	GTTTGTCCAA	2025
ACTCATCAAT	GTATCTTATC	ATGTCTGTAT	ACCGTCGACC	TCTAGCTAGA	GCTTGGCGTA	2085
ATCATGGTCA	TAGCTGTTTC	CTGTGTGAAA	TTGTTATCCG	CTCACAATTC	CACACAACAT	2145
ACGAGCCGGA	AGCATAAAGT	GTAAAGCCTG	GGGTGCCTAA	TGAGTGAGCT	AACTCACATT	2205
AATTGCGTTG	CGCTCACTGC	CCGCTTTCCA	GTCGGGAAAC	CTGTCTGTCC	AGCTGCATTA	2265
ATGAATCGGC	CAACGCGCGG	GGAGAGGCGG	TTTGCGTATT	GGGCGCTCTT	CCGCTTCCTC	2325
GCTCACTGAC	TCGCTGCGCT	CGGTCTGTTG	GCTGCGGCGA	GCGGTATCAG	CTCACTCAAA	2385
GGCGGTAATA	CGGTTATCCA	CAGAATCAGG	GGATAACGCA	GGAAAGAACA	TGTGAGCAAA	2445
AGGCCAGCAA	AAGGCCAGGA	ACCGTAAAAA	GGCCGCGTTG	CTGGCGTTTT	TCCATAGGCT	2505
CCGCCCCCCT	GACGAGCATC	ACAAAAATCG	ACGCTCAAGT	CAGAGGTGGC	GAAACCCGAC	2565
AGGACTATAA	AGATACCAGG	CGTTTCCCCC	TGGAAGCTCC	CTCGTGCGCT	CTCCTGTTCC	2625
GACCCTGCCG	CTTACCGGAT	ACCTGTCCGC	CTTCTCTCCCT	TCGGGAAGCG	TGGCGCTTTC	2685
TCAATGCTCA	CGCTGTAGGT	ATCTCAGTTC	GGTGTAGGTC	GTTGCTCCA	AGCTGGGCTG	2745
TGTGCACGAA	CCCCCGTTT	AGCCCGACCG	CTGCGCCTTA	TCCGGTAACT	ATCGTCTTGA	2805
GTCCAACCCG	GTAAGACACG	ACTTATCGCC	ACTGGCAGCA	GCCACTGGTA	ACAGGATTAG	2865
CAGAGCGAGG	TATGTAGGCG	GTGCTACAGA	GTTCTTGAAG	TGGTGGCCTA	ACTACGGCTA	2925
CACTAGAAGG	ACAGTATTTG	GTATCTGCGC	TCTGCTGAAG	CCAGTTACCT	TCGGAAAAAG	2985
AGTTGGTAGC	TCTTGATCCG	GCAAACAAAC	CACCGCTGGT	AGCGGTGGTT	TTTTTGTTTG	3045
CAAGCAGCAG	ATTACGCGCA	GAAAAAAAGG	ATCTCAAGAA	GATCCTTTGA	TCTTTTCTAC	3105
GGGGTCTGAC	GCTCAGTGGA	ACGAAAATC	ACGTTAAGGG	ATTTTGGTCA	TGAGATTATC	3165
AAAAAGGATC	TTCACCTAGA	TCCTTTTAAA	TTAAAAATGA	AGTTTTAAAT	CAATCTAAAG	3225
TATATATGAG	TAACTTGGT	CTGACAGTTA	CCAATGCTTA	ATCAGTGAGG	CACCTATCTC	3285
AGCGATCTGT	CTATTTCTGT	CATCCATAGT	TGCCTGACTC	CCCGTCGTGT	AGATAACTAC	3345
GATACGGGAG	GGCTTACCAT	CTGGCCCCAG	TGCTGCAATG	ATACCGCGAG	ACCCACGCTC	3405
ACCGGCTCCA	GATTTATCAG	CAATAAACCA	GCCAGCCGGA	AGGGCCGAGC	GCAGAAGTGG	3465
TCCTGCAACT	TTATCCGCCT	CCATCCAGTC	TATTAATTGT	TGCCGGGAAG	CTAGAGTAAG	3525
TAGTTCGCCA	GTTAATAGTT	TGCGCAACGT	TGTTGCCATT	GCTACAGGCA	TCGTGGTGTC	3585
ACGCTCGTCG	TTTGGTATGG	CTTCATTGAG	CTCCGGTTCC	CAACGATCAA	GGCGAGTTAC	3645
ATGATCCCCC	ATGTTGTGCA	AAAAAGCGGT	TAGCTCCTTC	GGTCCTCCGA	TCGTTGTCAG	3705
AAGTAAGTTG	GCCGCAGTGT	TATCACTCAT	GGTTATGGCA	GCACTGCATA	ATTCTCTTAC	3765

(2) INFORMATION FOR SEQ ID NO:34:

(A) LENGTH: 14 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

(2) INFORMATION FOR SEQ ID NO:35:

(A) LENGTH: 12 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

(2) INFORMATION FOR SEQ ID NO:36:

(A) LENGTH: 71 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Pro Ile Tyr Gly Val Pro Arg Tyr Ile Thr Tyr Gly Lys Trp Pro Ala  
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:37:

- (ii) MOLECULE TYPE: protein

Arg Gln  
1

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

Arg  
1

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

Met Ala Arg Leu Ala Leu Cys Pro Val His Asp Leu Met Gly Leu Ser  
1 5 10 15

Tyr Leu Ala Val His Leu Arg Ile Ser His Arg Tyr Tyr His Gly Asp  
20 25 30

Ala Val Leu Ala Val His Gln Trp Ala Trp Ile Ala Val  
35 40 45

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 10 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

(2) INFORMATION FOR SEQ ID NO:41:

- ```
(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:
```

(2) INFORMATION FOR SEQ ID NO:42:

- ```
(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:
```

(2) INFORMATION FOR SEQ ID NO:43:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Ala Cys Thr Val Gly Gly Leu Tyr Lys Gln Ser Ser Leu Ala Asn  
1 5 10 15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

(2) INFORMATION FOR SEQ ID NO:45:

(ii) MOLECULE TYPE: protein

Leu  
1

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

Cys His Leu Asn  
1